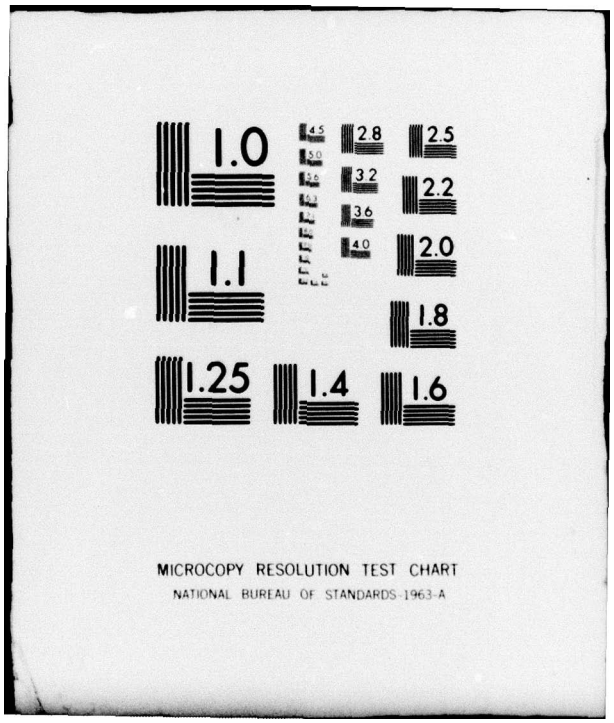
AD-A079 218 ARMY PERSONNEL RESEARCH OFFICE WASHINGTON DC F/G 5/10
SELF-TUITION AS A METHOD OF ADMINISTERING THE SEMANTIC TEST OF --ETC(U)
AUG 62 A G BAYROFF , E F HEERMANN

UNCLASSIFIED APRO-RM-62-3 NL

1 OF 1
AD-
A079218

END
DATE
FILMED
2-80
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Army Project Number
0J95-60-001

Input Quality c-12

(9) memo,
Research ~~Memorandum~~ 62-3

(6) SELF-TUITION AS A METHOD OF ADMINISTERING THE SEMANTIC TEST OF INTELLIGENCE, NQT-4,

(10) A. G. Bayroff, E. F. Heermann [redacted] A. A. Anderson

(14) APRO-RM-62-3

(13) 42

Submitted by

E. F. Fuchs
Chief, Military Selection Research Laboratory

(11) Aug [redacted] 62

040 650

SELF-TUITION AS A METHOD OF ADMINISTERING THE SEMANTIC TEST OF INTELLIGENCE, NQT-4

PURPOSE OF THE STUDY

The Nonlanguage Qualification Test, NQT-2 and -3, an experimental test developed for the Army ~~by Rulon (1959)~~, was designed to replace NQT-1. Thus, its main function in the screening process was to be assessment of the military trainability of inductees who have failed the AFQT and demonstrated a low level of literacy on the Verbal-Arithmetic subtest (VA). However, tryouts of NQT-2 and -3 revealed several difficulties in administration. The instructions were given in pantomine by the examiner and the test employed a format which involved frequent repetitions of a three-phase cycle:

1. Examiner explained and demonstrated the task.
2. Examinee completed non-scored practice items.
3. Examinee marked test items proper.

The pantomine instructions required that examiners be given special training. Since examinees found these pantomime instructions difficult to understand, one proctor had to be provided for every five examinees. The new format and length of the test resulted in a test administration lasting 2 1/4 to 2 1/2 hours.

To overcome these difficulties, NQT-2 and -3 were modified to permit self-tuition, by which examinees obtain immediate knowledge of the correctness of their responses without assistance from the proctor (Mundy, Tye, and Schenkel, 1956). Self-tuition was believed to aid the examinee in discovering the nature of the task without undue assistance from the proctor. After exploring various techniques for providing self-tuition (mechanical, electrical and chemical), a chemical reaction method was adopted. Pages of the test booklet, which also serve as the answer sheet, were chemically treated so that the examinee's mark--made with a special pen--was red for wrong alternatives and blue for correct alternatives. Only 20 pages from the 42 pages of NQT-2 were processed in this manner. The new test was designated NQT-4X.

Quite obviously the NQT-4X could not be made operational until its correlation with NQT-1 (which it was designed to replace) was determined and alternative methods of scoring had been evaluated. A secondary purpose of the study was to determine if the self-tuition feature of NQT-4X and the novel form of the items were valuable enough to warrant considering the incorporation of these features in other screening tests.

The present report describes analysis of data obtained in an exploratory administration of NQT-4X to determine the feasibility of the technique employed. If the results were favorable, a full-scale validation and standardization would be undertaken.

CONTENT AND FORMAT

Fundamentally, the NQT-4X is a test of the ability to learn the symbols of an artifical language (Table 1). In the first twelve pages of the test, nine symbols are defined pictorially. After each definition is a set of items which require the examinee to select from five pictures the one picture which correctly defines the symbol. The symbols defined in these twelve pages represent three nouns: boy, woman, dog; three intransitive verbs: running, lying, walking; and three transitive verbs: striking, pulling, chasing. The test items which follow (pages 13 and 14) test for the definition of all nine symbols. More complex items follow (pages 15 and 16) in which two symbols, a noun and verb, are put together so as to make a sentence; the examinee must select the correct picture to represent the meaning of the sentence. The items progress to three-symbol sentences (pages 17 and 18) and four-symbol sentences (pages 19 and 20). Since definitions of symbols needed to answer a set of items always reappear with the set of items, the test does not require the examinee to memorize the symbol definition; of course, some degree of memorization will facilitate the speed with which the test items are solved.

Since NQT-4X items are of the self-tuitional variety, verbal directions (not previously given) and one page of six demonstration items (the first page of the test booklet) were provided at the beginning of the test. No other guidance was offered throughout the test. One and one-half hours was allowed for taking the test.

In general, the nature of the items and the format represent an effort to construct a test of ability which is to a large extent free from the differential influence of past experience. No examinee could have had the opportunity to learn the meaning of the symbols prior to taking the test. Each examinee was given an equal number of practice trials in which to learn each symbol tested for. Following Rulon, this format is referred to as the semantic test format.

METHOD

The basic design of this study involved determining the correlation coefficients between varying combination of NQT-4X pages, scored by various formulae, and the currently operational NQT-1.

Table 1

CONTENT OF NQT-4X

Page	Picture Definition of New Symbol	Definitions Repeated	Symbol Tested For	No. Symbols In Lead For Items	No. Items
1	dog	none	dog	1	6
2	woman	none	woman	1	6
3	none	dog, woman	dog, woman	1	7
4	boy	none	boy	1	6
5	none	dog, woman, boy	dog, woman, boy	1	7
6	running	none	running	1	6
7	none	dog, woman boy, running	dog, woman boy, running	1	6
8	lying	none	lying	1	6
9	walking	none	walking	1	6
10	striking	none	striking	1	6
11	pulling	none	pulling	1	6
12	chasing	none	chasing	1	6
13,14	none	all previous	all previous	1	14
15,16	none	all previous	all previous	2-symbol sentences	14
17,18	none	all previous except running, lying, walking	all previous except running, lying, walking	3-symbol sentences	13
19,20	none	all previous	all previous	4-symbol sentences	13

SAMPLE

The appropriate population for evaluating NQT-4X would have been men who failed the AFQT and the VA subtest, that is, the men to whom NQT-1 is normally given. However, in such a population the range of AFQT scores is severely restricted. Since a somewhat larger range of scores on AFQT was desirable, a sample of 224 cases was selected in which AFQT percentile scores ranged from 1 to 19; the sample was not limited to VA subtest failures. During July and February 1959, the NQT-4X was administered to Selective Service registrants at the Atlanta, Newark, and New York AFES. Usable cases were divided into three subsamples:

Subsample A. AFQT percentile scores 10-19 (N = 35)

Subsample B. AFQT percentile scores 5-9 (N = 86)

Subsample C. AFQT percentile scores 1-4 (N = 103)

VARIABLES

As a basis for analysis to determine an appropriate scoring formula for NQT-4X items and to select NQT-4X pages which should be scored, several alternative scoring procedures were established and a representative set of pages was selected.

Scoring procedures. Traditional scoring formulas developed to correct for chance success were not applicable to NQT-4X items because an examinee was allowed to continue marking alternatives until the correct answer was achieved. The three scoring methods explored in the present analysis are described below:

1. Rights only. Credit was given only for those items which the examinee got right on the first trial. Items marked correctly after the first trial were not scored.

2. Total errors. Another procedure entailed counting the number of errors made per item in achieving a correct answer, and utilizing the total number of errors made as the examinee's score.

3. Corrected errors. In neither of the above procedures was allowance made for omitted items. To make such allowance, an error weight was assigned to omitted items with the assumption that omitted items, had they been attempted, would have been answered with chance success. On a chance basis, one-fifth of those attempting an item would make no errors, one-fifth would make one error, one-fifth would make two errors, one-fifth would make three errors, and one-fifth would make four errors. The median number of errors for a group answering the item on a chance basis would thus be 2, which was taken as the expected

number of errors for an omitted item. The formula for correcting errors to allow for omits is, then,

$E_c = E + 2 \times O$, where

E_c = corrected errors

E = total errors

O = number of omitted items

Pages selected for scoring. On the basis of the average p-value computed for each page in the total sample, the following pages of NQT-4X were selected as being the most meaningful for statistical analysis:

Pages	Average p value
5, 6, and 7 combined (symbol definition)	.73
13 and 14 combined (symbol definition)	.56
15 and 16 combined (Two-symbol sentences)	.32
17 and 18 combined (Three-symbol sentences)	.25
19 and 20 combined (Four-symbol sentences)	.22
5, 6, 7 and 13-20 combined	-

Since three scoring formulas were employed on each page, each of the page combinations yielded three scores: rights only, total errors, and corrected errors. Table 2 summarizes the experimental variables (page combinations and scoring formula) analyzed in this study.

Reference Variables. NQT-1 and VA subtest scores were obtained.

RESULTS

Table 2 gives the product moment correlation coefficients of the experimental variables with the reference variables. None of the variously scored page combinations yielded satisfactory coefficients with NQT-1, even though many of the coefficients were significantly different from zero. Accordingly, no further inquiry into the matter of selecting scoring formulae or page combinations to be employed with NQT-4X was considered appropriate. Further refinements of the NQT-4X as a replacement for NQT-1 would appear to be futile.

However, since NQT-4X represents a unique effort in test construction, the data were examined for any possible light on the potential usefulness of the self-tuition principle and the semantic format of the test. Because a control group was not included in the design, the present study did not permit a rigorous evaluation of the self-tuition principle. Previously published results provide the only baseline available for evaluating immediate knowledge of results as a test technique.

Previous research with the self-tuition principle yielded evidence implying that immediate knowledge of results given early in a series of items of the same type would lead to an increase in number of rights for the later items of the same series. Examination of the number of correct responses to a series of items of the same type failed to confirm this expectation. There were six test items for each of the eight symbols defined in the test (excluding the first page which was not scores). No rapid, abrupt, or even consistent improvement was observed from item to item (Table 3).

Why did immediate knowledge of results fail to confirm expectations derived from previous research? In the first place, it may be that the AFQT failures used in this sample did not profit from the type of knowledge of results given in the test. For a low level examinee, knowing that he is right or wrong may not be as important as knowing why he was right or wrong. In the second place, immediate knowledge of results may have failed to operate as expected because *nowhere* in the test was the examinee given a sufficient number of trials with any particular symbol. Immediate knowledge of results may not produce a noticeable effect when the number of trials is small. In the third place, the color difference between right and wrong responses was not sharply defined, although this difference became more sharply differentiated with the passage of time. The technical inadequacy of the color coding process may have contributed to failure of immediate knowledge of results. Finally, it is not inconceivable that the effectiveness of immediate knowledge of results is dependent upon the type of learning task on which it is applied. The items of the NQT-4X may constitute a type of learning task which is not facilitated by the application of immediate knowledge of results.

IMPLICATIONS FOR FUTURE RESEARCH

Future empirical evaluation of the principle of immediate knowledge of results should take cognizance of probable reasons for the failure of the principle in this study. Notably, if color coding is to be employed to convey knowledge of results, colors of the right and wrong marks on the answer sheet must be discriminably different and this difference must be immediately apparent to the examinee. NQT-4X did not meet this criteria, and should not be used in further evaluative studies. A more efficient means of providing immediate knowledge of results during test taking--a testing machine with knowledge of results programmed, for example--should be used instead.

Table 2

CORRELATION COEFFICIENTS OF EXPERIMENTAL VARIABLES WITH NQT-1 AND VA SUBTEST KEY

Content	Pages	Scoring Formula	NQT-1 (N = 190) r	VA (N = 121) r
Symbol definition	5,6,7 combined	Rights Only	.40*	.14
		Errors Only	-.35*	-.17
		Corrected Errors	-.37*	-.17
Symbol definition (all symbols previously presented)	13 and 14 combined	Rights Only	.44*	.17
		Errors Only	-.38*	-.20*
		Corrected Errors	-.41*	-.20*
2-symbol sentences	15 and 16 combined	Rights Only	.21*	.13
		Errors Only	-.28*	-.16
		Corrected Errors	-.31*	-.17
3-symbol sentences	17 and 18 combined	Rights Only	.19*	.26
		Errors Only	-.16*	-.19*
		Corrected Errors	-.20*	-.18*
4-symbol sentences	19 and 20 combined	Rights Only	.06	.27*
		Errors Only	-.03	-.27*
		Corrected Errors	-.11	-.30*
All Items	5,6,7, and 13 thru 20 combined	Rights Only	.38*	.25*
All Items	5,6,7, and 13 thru 20 combined	Errors Only	-.34*	-.28*
All Items	5,6,7, and 13 thru 20 combined	Corrected Errors	-.38*	-.28*

*Significant at $p < .05$

Table 3

PERCENTAGE INCREASE OR DECREASE OF SAMPLE GIVING CORRECT RESPONSES TO SEQUENCES OF SIMILAR ITEMS (N = 190)

	Page							
	2	4	6	8	9	10	11	12
Percentage Change From								
Item 1 to Item 2	8	6	4	4	-34	-16	-9	-14
Item 2 to Item 3	2	-11	4	-24	34	21	9	-6
Item 3 to Item 4	-7	-1	-4	19	-9	-08	-10	27
Item 4 to Item 5	9	16	12	6	19	16	3	-43
Item 5 to Item 6	-3	-4	-12	-3	-21	-21	4	29

In future research dealing with immediate knowledge of results, the nature of the test items--item difficulty, content, etc.--and the ability level of examinees should be systematically varied in order to provide results from which adequate generalizations can be drawn.

The semantic format of the NQT-4X must also be subjected to further experimentation before judgment of its usefulness can be rendered. One important variable in this experimentation would be the number of practice trials needed by individuals of different levels of ability to reach a successful solution of the test items. Individuals of low ability can be expected to require more practice trials than individuals of high ability. The test in its present form may not include enough practice trials so that low level individuals can perform satisfactorily on the test problems.

In summary, the NQT-4X should be abandoned as a potential screening device and as a research instrument for evaluating immediate knowledge of results. Future research evaluating immediate knowledge of results is contingent upon finding a technically more adequate method of providing knowledge of results. The research design should take into consideration the ability level of the experimental group and the nature of the test items, and should insure an adequate number of items of a certain type.

REFERENCES

Publications of U. S. Army Personnel Research Office, Office, Chief Research and Development.

1. Mundy, J. P., Tye, V. M., and Schenkel, K. F. Modification of the Nonlanguage Qualification Test to permit self-tuition. Research Memorandum 56-6. February 1956.

2. Rulon, P. J. Development of a nonverbal classification test. Technical Research Report 877. November 1950.

3. Rulon, P. J., and Schweiker, R. F. Validation of a non-verbal test of military trainability. Technical Research Report 1076. June 1953.